J. Fluid Mech. (2019), *vol.* 861, *pp.* 886–900.

doi:10.1017/jfm.2018.898

Model parameter estimation using coherent structure colouring

Kristy L. Schlueter-Kuck[1] **and John O. Dabiri**[1,2,†]

[1]Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA

[2]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA

(Received 6 April 2018; revised 13 September 2018; accepted 4 November 2018; first published online 28 December 2018)

Lagrangian data assimilation is a complex problem in oceanic and atmospheric modelling. Tracking drifters in large-scale geophysical flows can involve uncertainty in drifter location, complex inertial effects and other factors which make comparing them to simulated Lagrangian trajectories from numerical models extremely challenging. Temporal and spatial discretisation, factors necessary in modelling large scale flows, also contribute to separation between real and simulated drifter trajectories. The chaotic advection inherent in these turbulent flows tends to separate even closely spaced tracer particles, making error metrics based solely on drifter displacements unsuitable for estimating model parameters. We propose to instead use error in the coherent structure colouring (CSC) field to assess model skill. The CSC field provides a spatial representation of the underlying coherent patterns in the flow, and we show that it is a more robust metric for assessing model accuracy. Through the use of two test cases, one considering spatial uncertainty in particle initialisation, and one examining the influence of stochastic error along a trajectory and temporal discretisation, we show that error in the coherent structure colouring field can be used to accurately determine single or multiple simultaneously unknown model parameters, whereas a conventional error metric based on error in drifter displacement fails. Because the CSC field enhances the difference in error between correct and incorrect model parameters, error minima in model parameter sweeps become more distinct. The effectiveness and robustness of this method for single and multi-parameter estimation in analytical flows suggest that Lagrangian data assimilation for real oceanic and atmospheric models would benefit from a similar approach.

Key words: chaotic advection, geophysical and geological flows, nonlinear dynamical systems

1. Introduction

Models of oceanic and atmospheric geophysical flows are of great importance in weather prediction and nowcasting (Kalnay 2002), understanding the effects of

† Email address for correspondence: jodabiri@stanford.edu

natural disasters such as tsunamis (Ioualalen *et al.* 2007; Yamazaki *et al.* 2011) and anthropogenic disasters such as oil spills (Mariano *et al.* 2011) and quantifying transport of nutrients and passive tracers such as heat and salt in the ocean (Böning & Hermann 1994). Due to the large length scales of such flows, they are frequently directly measured by sparse sensors that approximately follow the flow, including ocean drifters and weather balloons. These sensors provide valuable data on flow velocities, temperature, density and other parameters that can be used to validate and tune the numerous parameters in the complex models developed to simulate these flows. However, the integration of the real Lagrangian flow data into typical Eulerian model descriptions is complicated by the nonlinear relationship between the underlying flow velocities and the paths taken by individual parcels of fluid or fluid tracers. In flows with significant chaotic advection, even small uncertainties in spatial or temporal initial conditions can lead to a large divergence in the trajectories taken by tracers that are initially closely spaced (Guckenheimer & Holmes 1983). Further, real drifters experience effects such as added mass, Basset forces and windage that influence their trajectories (Putman & He 2013; Olivieri *et al.* 2014). Because oceanic and atmospheric models frequently use error in physical positions of simulated Lagrangian particles compared to real drifters in the flow to assess the accuracy of the models (Apte, Jones & Stuart 2008), the combination of these factors can complicate parameter determination leading to large errors even for physically correct models. All of the factors noted previously need to be accounted for when comparing real drifters to simulated drifters for the purposes of Lagranginan data assimilation. Often even a small uncertainty in these factors can lead to large departures of the simulated trajectories from the real drifters. It is, therefore, useful to make sure any method for trajectory comparison is robust to reasonable levels of uncertainty.

This issue necessitates the development of error metrics that are robust to the effects of chaotic advection and uncertainties in drifter position. Recent advances in coherent pattern identification provide a potential solution. One goal of coherent pattern (or coherent structure) identification is to gain insight into transport barriers in flows. This can be accomplished by identifying lines or surfaces of maximal separation of initially closely spaced particles, as in finite time Lyapunov exponent (FTLE) analysis (Haller 2000; Shadden, Lekien & Marsden 2005), by clustering trajectories using machine learning algorithms (Froyland & Padberg-Gehle 2015), identifying groups of trajectories that remain spatially compact (Hadjighasem *et al.* 2016), bounding intertwining lines in a hybrid spatial/temporal parameter space (Thiffeault 2010) or clustering trajectories based on their relative dissimilarities (Schlueter-Kuck & Dabiri 2017*a*). With the exclusion of FTLE analysis, these techniques have been developed to be compatible with sparsely initialised tracer distributions, making them well suited for analysis of ocean drifters and weather balloons. The underlying flow structure represented by these coherent patterns can potentially provide an effective metric for quantifying error in models.

Recently, one innovative study combined these ideas of coherent structure identification and model parameter estimation for Lagrangian data assimilation. In this study, the authors used principal component analysis (PCA) of the evolution of the horizontal position of a set of drifters initialised around the centre of two gyre cores in an analytical flow field (Maclean, Santitissadeekorn & Jones 2017). They successfully identified model parameters when a random component was added to the advected drifter location at every time step. However, this study did not address several factors that complicate the use of global positioning system (GPS)-enabled drifters to map ocean currents. Namely, while the positions of drifters may be

known with high accuracy, even minimal error in the initial location of the drifters can significantly impact the drifter trajectory. Furthermore, due to the difficulty in simultaneously deploying drifters, irregular or random initialisation of drifters must be accounted for. Here, we propose a method of parameter estimation based on coherent structure colouring, which uses the dissimilarity in the drifter trajectories to identify the underlying coherent flow patterns. This method has been shown to be robust to measurement uncertainty and data loss, is effective even in instances of sparse data (approximately 300 drifters tracked in a two-dimensional flow) and is effective at identifying coherent sets of trajectories (Schlueter-Kuck & Dabiri 2017*b*).

This analysis is focused on two specific test cases. In the first test case, the effects of error in spatial initialisation are examined. In the second case, effects on parameter estimation due to temporal discretisation, error in spatial initialisation and stochastic error along the drifter trajectories (meant to represent any number of inertial effects and/or spatial discretisation), are quantified. Section 2 details the computational setup of the analysis and § 3 presents results of the analysis. Conclusions and future directions are discussed in § 4.

2. Methods

2.1. *Test case 1: drifter position uncertainty*

The first test case examined here seeks to quantify the effect of uncertainty in the initial location of Lagrangian drifters on the parameter estimation process. Because chaotic advection in fluid flows tends to separate even closely spaced fluid particles over sufficiently long periods of time, even a small uncertainty in spatial initialisation can lead to large particle displacement errors (e.g., Putman & He 2013). For this case, the analytical quadruple gyre, governed by (2.1)–(2.3), is used to model the underlying flow:

$$u_x = \frac{\mathrm{d}x}{\mathrm{d}t} = -\pi\alpha \sin(\pi f)\cos(\pi y), \tag{2.1}$$

$$u_y = \frac{\mathrm{d}y}{\mathrm{d}t} = \pi\alpha \cos(\pi f)\sin(\pi y)(2ax + b), \tag{2.2}$$

where x and y are the spatial coordinates, t is time and

$$a = \epsilon \sin(\omega t), \quad b = 1 - 2\epsilon \sin(\omega t), \quad f = ax^2 + bx. \tag{2.3a–c}$$

Parameter values of $\alpha = 0.1$, $\epsilon = 0.1$ and $\omega = 2\pi/10$ are used, giving the flow a periodic east-west oscillation.

To build a set of 'real' trajectories, analogous to an array of drifter positions at discrete moments in time in an oceanic or atmospheric flow, a set of 500 drifters is randomly initialised in the domain $x = [0, 2]$, $y = [-1, 1]$. These drifters are advected using a fifth-order Runge–Kutta integration scheme with a relative error tolerance of 10^{-6} and an absolute error tolerance of 10^{-9}. These parameters are used for calculation of all 'real' trajectories in this study. The 'simulated' particles are initialised near the initial locations of the real trajectories, with an offset of 0.18 in a random direction, corresponding to 9 % of the length of the domain. For the single-parameter analysis, the parameter ϵ in (2.3) is assumed to be unknown, and a set of simulated trajectories is created for 201 equally spaced values of ϵ in the range $\epsilon = [0, 0.4]$. For each value of ϵ, 10 independent simulations with different initial conditions are run in order to examine the repeatability of the results. For the multiple-parameter analysis, α, ω and

ϵ in (2.1)–(2.3) are assumed to be unknown, and a set of simulated trajectories is created for all combinations of 21 equally spaced values of α in the range $\alpha = [0, 0.4]$, 21 equally spaced values of ω in the range $\omega = [0, \pi]$ and 21 equally spaced values of ϵ in the range $\epsilon = [0, 0.4]$. Each unique combination of parameters is tested using 10 independent simulations, resulting in 92 610 simulations of 500 particles each for the multi-parameter study.

2.2. *Test case 2: temporal discretisation in model and stochastic position error along trajectories*

The second test case in this study examines the combined effects of model temporal discretisation and stochastic error on parameter estimation. Because there are many factors that influence the divergence of simulated drifter trajectories in models from the real, observed drifter trajectories, it is critical that any parameter estimation scheme be able to account for the complex interaction of several of these factors. This test utilises the analytical Bickley jet flow, frequently used as a model of zonal atmospheric currents (Rypina, Brown & Beron-Vera 2007), which is defined by the stream function $\psi = \psi_0 + \psi_1$, where

$$\psi_0 = c_3 y - UL \tanh\left(y/L\right), \tag{2.4}$$

$$\psi_1 = UL \operatorname{sech}^2\left(y/L\right) \sum_{n=1}^{3} \epsilon_n \cos\left(k_n \left(x - \sigma_n t\right)\right). \tag{2.5}$$

For this analysis, $U = 62.66$ m s^{-1}, $L = 1770$ km, $k_n = 2n/r_0$, $c = [0.1446U, 0.205U, 0.461U]$, $\sigma = c - c(3)$, and $\epsilon = [0.0075, 0.15, 0.3]$ and the flow is computed on the interval $x = [0, 20 \times 10^6]$ m, $y = [-3 \times 10^6, 3 \times 10^6]$ m over the time interval $t = [0, 80]$ days.

For this test, the 'real' trajectories are calculated by initialising 500 particles in the domain and advecting them using a fifth-order Runge–Kutta scheme, with the same error tolerances as noted previously. The drifters are advected over 80 days according to (2.4) and (2.5).

For this analysis, the 'simulated' particles are initialised near the initial locations of the real drifters, offset by 60 km (0.3 % of the horizontal spatial domain) in a random direction. Although 60 km is more uncertainty than could be expected from GPS data alone, this value is meant to capture some of the effect of spatial discretisation in the model as well, which is not directly considered. A random drifter initialised in the Bickley jet flow at $t = 0$ with a spatial resolution of 100 km will, on average, accumulate 60 km of error when compared to a 'real' drifter after one 30-minute time step if linear interpolation between grid points is used to estimate velocity. The particles are advected using Euler integration with a temporal discretisation of 30 min. At every time step, a stochastic error with a standard deviation of 60 m is added to the drifter position. For this analysis, the parameter $\epsilon(3)$ is assumed to be unknown, and a set of simulated trajectories is created for 201 equally spaced values of $\epsilon(3)$ in the range $\epsilon(3) = [0, 1]$.

2.3. *Coherent structure colouring and error quantification*

In each test case above, the underlying coherent flow pattern is quantified using the coherent structure colouring (CSC) vector calculated for the 'real' trajectories and each set of simulated trajectories (Schlueter-Kuck & Dabiri 2017*a*). In the CSC

algorithm, dissimilarity between two particle trajectories is numerically represented using a weighted adjacency matrix $\boldsymbol{A}$, where a_{ij} contains the weight of the edge connecting particle i and particle j:

$$a_{ij} = \frac{1}{\overline{r_{ij}} T^{1/2}} \left[\sum_{k=0}^{T-1} (\overline{r_{ij}} - r_{ij}(t_k))^2 \right]^{1/2}, \tag{2.6}$$

where $r_{ij}(t_k)$ is the distance between two particles i and j at time t_k and $\overline{r_{ij}}$ is the average distance between the two fluid particle trajectories. Conceptually, a_{ij} quantifies the standard deviation of the distance between particle trajectories normalised by their average spacing. The corresponding generalised eigenvalue problem that quantifies the difference between dissimilar particles is

$$\boldsymbol{L}X = \lambda \boldsymbol{D}X, \tag{2.7}$$

where

$$d_{ij} = \begin{cases} 0, & i \neq j \\ \sum_{k=1}^{N} a_{ik}, & i = j, \end{cases} \tag{2.8}$$

and $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ is the graph Laplacian. In order to maximise the differences between dissimilar particles, $X_1 = X$ is the eigenvector associated with the maximum eigenvalue, λ_1, of this problem, under the constraint that $X'\boldsymbol{D}X$ remains finite. Each element of X_1 assigns that value of CSC to the corresponding fluid particle. This vector can be visualised in the spatial domain by mapping each element of the CSC vector to the initial particle location and interpolating between particles to obtain a contour field.

Two error metrics are examined for each test case. The first metric, the average particle displacement error, is defined by

$$E_{disp} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |\boldsymbol{x}_{real}^{i,j} - \boldsymbol{x}_{sim}^{i,j}|, \tag{2.9}$$

where N is the number of drifters tracked, T is the number of time steps and $\boldsymbol{x}_{real}^{i,j}$ and $\boldsymbol{x}_{sim}^{i,j}$ are the spatial locations of the ith real drifter and simulated drifter, respectively, at the jth time step.

The error in the CSC field is calculated as follows:

$$E_{CSC_f} = \frac{1}{M} \sum_{i=1}^{M} \left| \mathrm{CSC}_{f,real}^{i} - \mathrm{CSC}_{f,sim}^{i} \right|, \tag{2.10}$$

where $\mathrm{CSC}_{f,real}^{i}$ and $\mathrm{CSC}_{f,sim}^{i}$ are the CSC field values of the ith Cartesian grid node for the set of real and simulated sets of drifters, respectively, and M is the number of Cartesian grid nodes in the CSC field. It is important to note that the error in the interpolated CSC field is found to be more effective for parameter estimation than the CSC vector itself, as it is more robust to individual particle position. This point will be expanded upon in the results section. The CSC field is calculated by interpolating the CSC vector onto a grid of spatial locations using triangulation-based linear interpolation, and M is the product of the number of discrete horizontal locations and

the number of discrete vertical locations. The error based on the CSC field is found to be insensitive to the interpolation grid coarseness, as long as the average spacing between grid elements is smaller than the average spacing between particles tracked. The CSC vector error, shown in the results section for comparison, is quantified as follows:

$$E_{CSC_v} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{CSC}^i_{v,real} - \mathrm{CSC}^i_{v,sim}\right|, \tag{2.11}$$

where $\mathrm{CSC}^i_{v,real}$ and $\mathrm{CSC}^i_{v,sim}$ are the CSC vector values of the ith drifter for the set of real and simulated sets of drifters, respectively, and N is the total number of drifters tracked.

3. Results

3.1. *Test case 1: drifter position uncertainty*

The first test case examined the effect of error in initial drifter position on the ability to accurately determine the parameter ϵ, the magnitude of the periodic horizontal oscillation in the quadruple gyre flow. Figure 1 examines the 'real' and simulated trajectories (with the correct value of the unknown parameter, $\epsilon = 0.1$ but a slightly errant initial position) for two initial positions, one within the lower left coherent vortex, and the second near the transport barrier located at the intersection of the four quadrants. It is clear that by the end of the simulated time interval of four horizontal oscillation cycles, chaotic advection in the flow has strongly separated the real drifter located near the centre of the domain (i.e., figure 1(*b*), black trajectory) from most of the simulated trajectories of the same drifter. In fact, there are simulated trajectories that end up in each of the four quadrants. This separation is due to the error in initial position resulting in the particles being initialised in a different quadrant of the flow. Subsequently, the transport barrier separating the 'real' particle from the corresponding simulated particles leads to an exponentially large spatial separation in time. In contrast, for the drifter initialised at the centre of the lower left vortex, the simulated trajectories remain compact, as there are no transport barriers separating the 'real' drifter initial position from the simulated initial positions.

Figure 2 shows the CSC fields for the real drifter trajectories (*a*), the simulated trajectories with the correct value of epsilon (*b*) and the simulated trajectories with incorrect values of $\epsilon = 0$ (*c*) and $\epsilon = 0.4$ (*d*). For the simulated case with the correct value of epsilon, the CSC field highlights the largest kinematic dissimilarity in the flow between the particles that remain in the gyre cores (with high CSC values) and those particles that switch quadrants during the prescribed time interval (with low CSC values). By comparing the CSC fields for the real and correctly simulated sets, it is evident that the CSC field is robust to chaotic advection of individual drifters. This is because the Lagrangian drifters are used as landmarks for interpolating the underlying coherent patterns in the flow, but do not individually dictate these patterns. It should be noted that the magnitude of CSC vector and field values is dependent on the number of particles used for analysis. Thus, when assessing relative error values, the number of particles tracked must be held constant. Furthermore, because the CSC vector is an eigenvector associated with the generalised eigenvalue problem given by (2.8), both the calculated CSC vector and its negative will need to be considered when comparing the real and simulated drifter sets, and the minimum error kept.

Because of the factors noted above, two independently initialised sets of particles should have the same CSC field, up to error on the order of particle separation

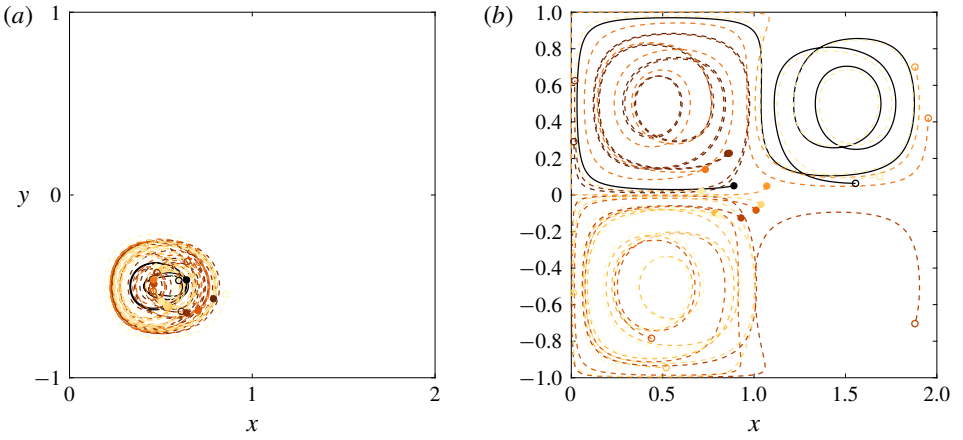

FIGURE 1. (Colour online) Sample drifters in quadruple gyre flow. Solid black lines indicate 'real' drifter trajectories and dashed lines of other colours indicate simulated drifter trajectories. Closed circles correspond to drifter initial locations at $t_0 = 2.5$ and open circles to drifter final locations at $t_f = 42.5$. (*a*) Drifter initialised in the centre of the lower left coherent vortex. (*b*) Drifter initialised near the intersection of the four quadrants.

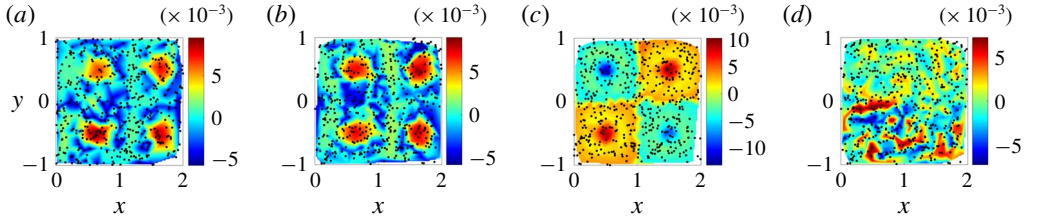

FIGURE 2. (Colour online) CSC contours for quadruple gyre overlaid with dots indicating initial positions of 500 drifters for $t = [2.5, 42.5]$. (*a*) 'Real' trajectories. (*b*) Simulated trajectories with errant initial position but correct value of $\epsilon = 0.1$. (*c*) Simulated trajectories with errant initial position and with incorrect value of $\epsilon = 0$. (*d*) Simulated trajectories with errant initial position and with incorrect value of $\epsilon = 0.4$.

distance, regardless of how different the spatial initialisations are. The CSC field for the incorrect value of $\epsilon = 0$ clearly has a different underlying flow structure, where the largest kinematic dissimilarity is between the quadrants that have counterclockwise rotation and those with clockwise rotation of fluid. Similarly, the CSC field for the incorrect value of $\epsilon = 0.4$ also has a flow structure distinct from the 'real' flow; the magnitude of the left-right oscillation is strong enough to eliminate the coherent vortices in each quadrant, and the flow mixes chaotically. This example highlights why using the error in the interpolated CSC field is more robust than using the error in the CSC vector. A drifter just outside the boundary of a coherent structure will be assigned a different CSC value than a drifter just inside the boundary, but the location of the boundary itself in the CSC field will be the same regardless of the location at which that particular drifter is initialised. By moving away from an individual trajectory analysis, and towards a field-based analysis of the underlying flow, correct model parameters can be identified unambiguously.

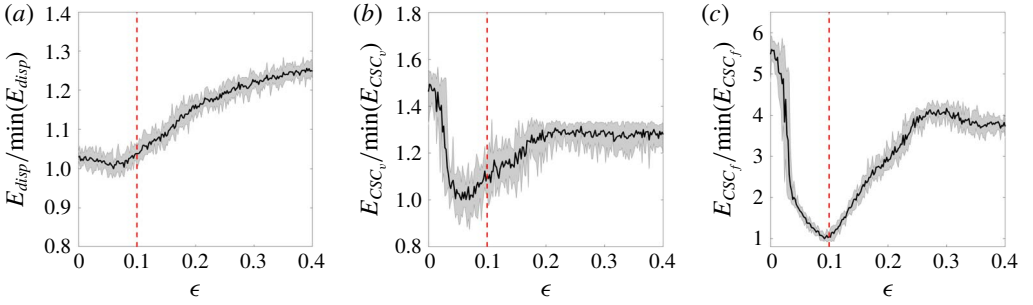

FIGURE 3. (Colour online) Normalised error versus candidate parameter ϵ. Error metrics are normalised by the lowest error over the range of ϵ tested. Solid black line indicates the error averaged over 10 individual sweeps and shaded region bounds the range of errors seen for an individual sweep. Red dotted line indicates the correct value for the parameter ϵ. (*a*) Particle displacement error; i.e., (2.9). (*b*) CSC vector error; i.e., (2.11). (*c*) CSC field error; i.e., (2.10).

The conventional particle displacement error and the CSC field error are compared in figure 3 as a function of the selected value of ϵ for the simulated particle set. While the CSC error is minimised at the correct value of $\epsilon = 0.1$, the particle displacement error and CSC vector error are not. Hence, model parameter estimation using the CSC field recovers the correct parameter value despite the potentially confounding effects of initial position uncertainty. The conventional error metric fails under the same conditions. It is also interesting to note that the CSC field error has a larger difference in error between the correct and incorrect states, with the error at $\epsilon = 0$ exceeding five times the error for $\epsilon = 0.1$. In contrast, the particle displacement error varies less than 30 % from the minimum error over the range of parameters tested. This means that not only does the CSC field aid in accurately predicting the correct model parameter, but it amplifies the error between correct and incorrect parameter values.

The purpose of the multi-parameter analysis is to evaluate the effectiveness of this method for determining more than one unknown parameter simultaneously, a problem that is frequently encountered in oceanic and atmospheric models. An analysis of the full parameter space, while not computationally efficient, is used here for completeness and visualisation purposes. It is beyond the scope of this work to investigate various optimisation techniques for determining the global error minimum, although any viable techniques would need to take into account the noisiness of the error signal as well as its non-convexity. These factors make common approaches such as gradient descent optimisation unfeasible.

Figure 4 shows the average particle displacement error (*a*) and average CSC field error (*b*) as a function of the three unknown parameters, α, ω and ϵ, averaging the error produced for each unique combination of parameters over 10 independent simulations with different initial offsets for each drifter. Data for only five of the 21 values of ϵ is displayed for simplicity. The location of the minimum in the error is highlighted by the white circles for each case. It is clear that both the particle displacement error and the CSC field error exhibit global minima in approximately the same region of the parameter space. The error minima are examined more closely in table 1, which shows the parameter values where the displacement, CSC vector and CSC field error are minimised for each of the individual 10 parameter sweeps as well

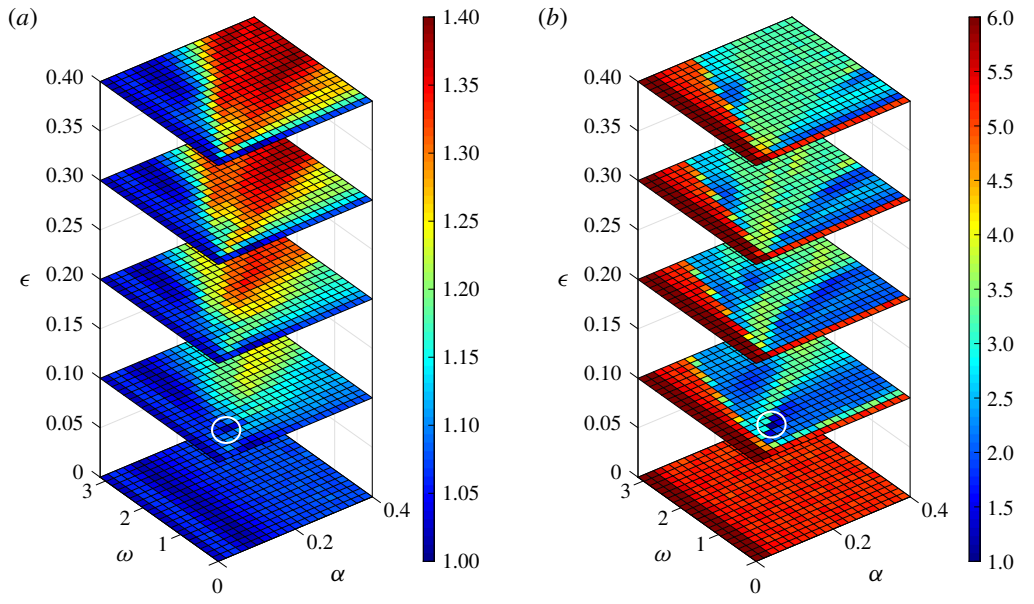

FIGURE 4. (Colour online) Normalised error metrics averaged over 10 individual three-parameter sweeps in α, ω, and ϵ. Error metrics are normalised by the lowest error over the range of all three parameters tested. (*a*) Displacement error for selected ϵ-slices. (*b*) CSC field error for selected ϵ-slices. White circles indicate the approximate location of the global minimum.

Error metric	E_{disp}	E_{CSC_v}	E_{CSC_f}
Run no.			
Average	(0.10, $2\pi/10$, 0.08)	(0.18, $5\pi/20$, 0.16)	(0.10, $2\pi/10$, 0.10)
1	(0.08, $2\pi/10$, 0.08)	(0.18, $5\pi/20$, 0.08)	(0.10, $2\pi/10$, 0.10)
2	(0.10, 0 , 0.10)	(0.38 $7\pi/20$, 0.30)	(0.10, $2\pi/10$, 0.10)
3	(0.10, $2\pi/10$, 0.08)	(0.28, $5\pi/20$, 0.24)	(0.10, $2\pi/10$, 0.10)
4	(0.06, $7\pi/20$, 0.04)	(0.26, $5\pi/20$, 0.28)	(0.10, $2\pi/10$, 0.10)
5	(0.08, 0 , 0.34)	(0.22, $5\pi/20$, 0.18)	(0.10, $2\pi/10$, 0.10)
6	(0.08, $19\pi/20$, 0.12)	(0.26, $3\pi/10$, 0.24)	(0.10, $2\pi/10$, 0.10)
7	(0.10, $2\pi/10$, 0.08)	(0.20, $5\pi/20$, 0.14)	(0.10, $2\pi/10$, 0.10)
8	(0.08, π, 0.26)	(0.18, $2\pi/10$, 0.20)	(0.10, $2\pi/10$, 0.10)
9	(0.10, $2\pi/10$, 0.08)	(0.36, $7\pi/20$, 0.22)	(0.10, $2\pi/10$, 0.10)
10	(0.10, $2\pi/10$, 0.06)	(0.36, $3\pi/10$, 0.24)	(0.10, $2\pi/10$, 0.10)

TABLE 1. Global minima, (α_{min}, ω_{min}, ϵ_{min}), for each of 10 individual three-parameter sweeps, along with the global minimum in the error averaged over all 10 sweeps.

as for the averaged error field for each metric. These data show that while averaging data from 10 sweeps results in a nearly correct identification of all three unknown parameters for both the displacement error and the CSC field error, individual sweeps of the displacement error result in wildly incorrect parameter identification, while the CSC field error results in a global minimum at the correct parameter-space location for every single individual parameter sweep. The tenfold-reduction in the

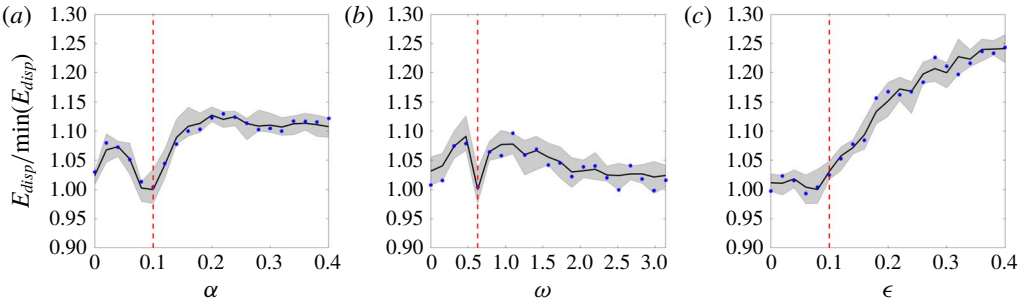

FIGURE 5. (Colour online) Normalised displacement error slices through the three-parameter sweep. Solid black line indicates the error averaged over 10 individual sweeps and shaded region bounds the range of errors seen for an individual sweep. Blue stars indicate error for a single representative sweep. Dotted red lines indicate the correct value of the parameter. (*a*) Slice at $\omega = 2\pi/10$, $\epsilon = 0.08$. (*b*) Slice at $\alpha = 0.10$, $\epsilon = 0.08$. (*c*) Slice at $\alpha = 0.10$, $\omega = 2\pi/10$.

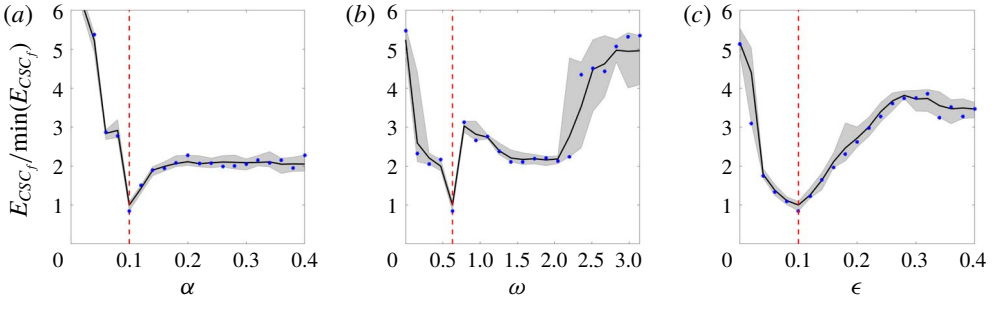

FIGURE 6. (Colour online) CSC field error slices through the three-parameter sweep. Solid black line indicates the error averaged over 10 individual sweeps and shaded region bounds the range of errors seen for an individual sweep. Blue stars indicate error for a single representative sweep. Dotted red lines indicate the correct value of the parameter. (*a*) Slice at $\omega = 2\pi/10$, $\epsilon = 0.10$. (*b*) Slice at $\alpha = 0.10$, $\epsilon = 0.10$. (*c*) Slice at $\alpha = 0.10$, $\omega = 2\pi/10$.

number of simulations needed for correct parameter identification using the CSC field error is potentially extremely beneficial for complex and computationally expensive simulations of oceanic and atmospheric flows, where large simulation ensembles are currently required in practice (Kalnay 2002). Figure 5 shows slices of the error field for the displacement error at the parameter values where the global minimum is identified, in this case at $(\alpha_{min}, \omega_{min}, \epsilon_{min}) = (0.10, 2\pi/10, 0.08)$. Data from one of the individual sweeps is plotted along with the average of all 10 sweeps and the range of error values resulting for each parameter combination. Figure 6 shows comparable data for the CSC field error, with slices taken at $(\alpha_{min}, \omega_{min}, \epsilon_{min}) = (0.10, 2\pi/10, 0.10)$. It is clear from these plots that there is less variation among the 10 sweeps for the CSC field error, especially in regions of the parameter space close to the global minimum. This feature of the CSC field error allows for correct identification of all of the unknown parameters simultaneously, without the need for an ensemble of multiple simulation iterations.

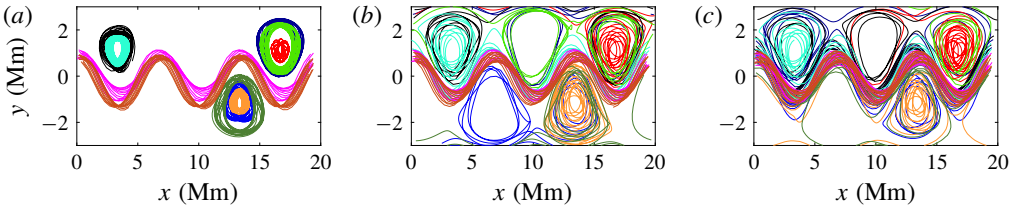

FIGURE 7. (Colour online) Ten drifters in the Bickley jet flow. (*a*) 'Real' trajectories. (*b*) Trajectories with temporal discretisation, no stochastic position error and correct value of $\epsilon(3) = 0.3$. (*c*) Trajectories with temporal discretisation, stochastic position error and correct value of $\epsilon(3) = 0.3$.

3.2. *Test case 2: temporal discretisation in model and stochastic position error along trajectories*

The second test case uses the Bickley jet flow to highlight the effect that temporal discretisation and position error, both initially and along the drifter trajectory, have on the process of determining model parameters accurately. Figure 7 shows a sample of 10 of the 500 total trajectories for the 'real' set of drifters (*a*); the simulated set with the correct value of $\epsilon(3) = 0.3$, with temporal discretisation but no position error (*b*); and the simulated set with the correct value of $\epsilon(3) = 0.3$, where both temporal discretisation and position error are included (*c*). Temporal discretisation and stochastic position error both tend to push drifters near the centres of the gyre cores towards the edge of the gyres or even out completely into the background flow, although in all three cases highlighted in figure 7, the zonal meandering jet and gyre cores remain distinct. It is clear that both temporal discretisation and stochastic position errors will act to degrade the effectiveness of the error analysis, and if pushed to extremes, will render any error metric ineffective at parameter estimation. The question, in this case, is whether the minimum in the displacement error and/or CSC field error are robust to reasonable levels of these effects. One potentially useful technique in countering the effects in this study is selective drifter placement. Here, the set of 'real' drifters are initialised near the centre of the coherent structures in the flow, as dictated by the ridges in the FTLE field. Figure 8 shows the selected 'real' drifter initial positions, with the FTLE field in the background. Practically, it might be challenging to know where to initialise real drifters in an oceanic or atmospheric flow, but this could potentially be accomplished using iterative deployments: the first to determine drifter placement for the second, and the second set of drifters used to tune model parameters. Figure 9 shows the CSC field for the real set of particles (*a*), the simulated set of particles with temporal discretisation and stochastic position error with the correct value of $\epsilon(3) = 0.3$ (*b*) and the simulated set of particles with an incorrect value of $\epsilon(3) = 0$ (*c*). For both the real set of trajectories and the simulated set with the correct value of $\epsilon(3)$, the flow is dominated by the meandering jet and the flanking vortices. Despite the addition of error in trajectory calculation, the CSC field for the simulated sets of trajectories still exhibits distinct coherence. For the incorrect values of $\epsilon(3)$, the jet has a much wider vertical extent, and only portions of the flanking vortices are seeded, while the extent of the jet for the correct value of $\epsilon(3)$ is similar to that of the real flow. The particle displacement error, CSC vector error, and CSC field error for the full range of tested parameters are shown in figure 10. As with the previous study, each parameter value is tested using 10 individual simulations, and the average and range of error for each parameter

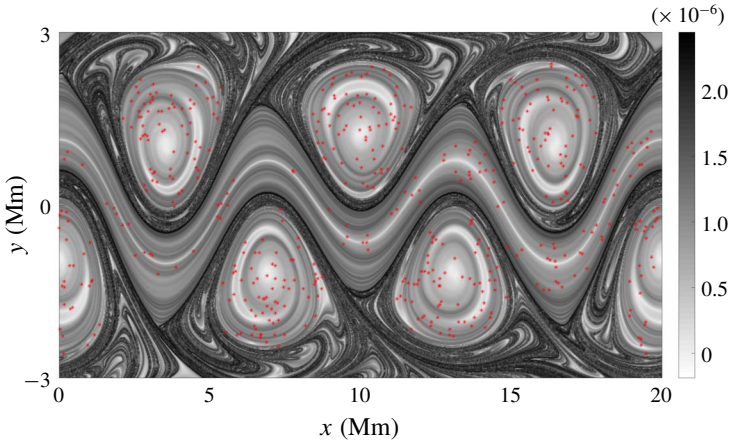

FIGURE 8. (Colour online) Initial positions for the 'real' drifters in the Bickley jet study, indicated by the red stars. Background shows the FTLE field calculated over the time interval $t = [0, 3456 \times 10^3]$ s.

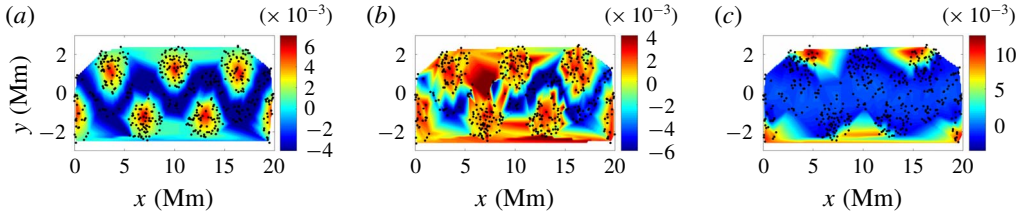

FIGURE 9. (Colour online) CSC contours for the Bickley jet flow overlaid with dots indicating initial positions of 500 drifters for $t = [0, 6912 \times 10^3]$ s. (*a*) 'Real' trajectories. (*b*) Simulated trajectories with temporal discretisation and stochastic position error, but correct value of $\epsilon(3) = 0.3$. (*c*) Simulated trajectories with temporal discretisation, stochastic position error and incorrect value of $\epsilon(3) = 0$.

value is shown in the figure, along with the correct value of $\epsilon(3) = 0.3$. The same trends seen in test case 1 are also present here. The minimum error in the CSC field is very close to the correct value of $\epsilon(3)$, with identified values of $\epsilon(3)$ ranging from 15 % below the real value to 9 % above. However, the error metric based on particle displacement identifies values of $\epsilon(3)$ ranging from 7 % to 20 % below the real value. In this case, the CSC vector error comes close to estimating the unknown parameter correctly, in contrast with the study focusing solely on initial particle error. Additionally, the CSC error metric enhances the difference in error between correct and incorrect parameter values over particle location error, as with the previous study (i.e., compare the vertical axes of the panels in figure 10).

It is useful to understand the effects of using a random drifter distribution instead of a set of drifters initialised inside the coherent structures in the flow. Figure 11 shows the CSC field for simulated set of trajectories with a correct value of $\epsilon(3)$ (panel *a*) and the CSC field error for a range of epsilon values, again using a random initial particle distribution and otherwise identical parameters to the study discussed in this section. Due to the lack of particles remaining in the vortex cores, the CSC analysis

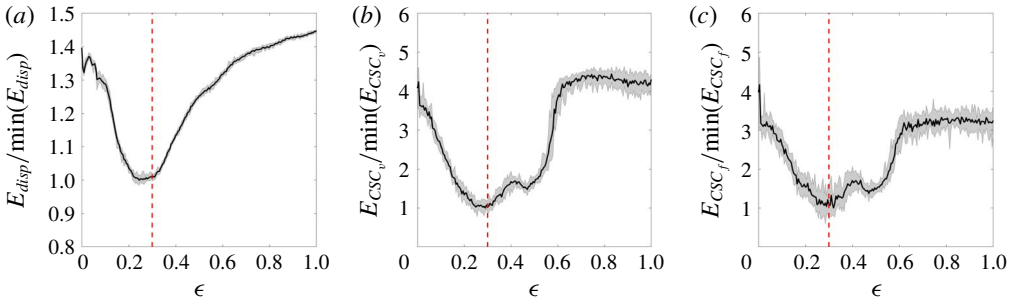

FIGURE 10. (Colour online) Normalised error versus candidate parameter $\epsilon(3)$. Error metrics are normalised by the lowest error over the range of $\epsilon(3)$ tested. Solid black line indicates the error averaged over 10 individual sweeps, and shaded region bounds the range of errors seen for an individual sweep. Red dotted line indicates the correct value for the parameter $\epsilon(3)$. (*a*) Particle displacement error; i.e., (2.9) (*b*) CSC vector error; i.e., (2.11) (*c*) CSC field error; i.e., (2.10).

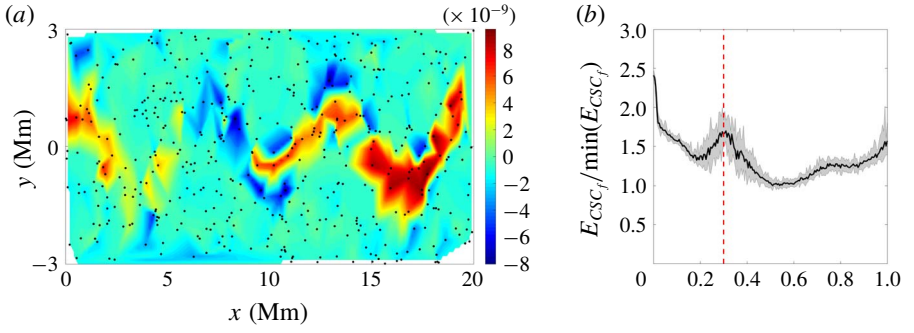

FIGURE 11. (Colour online) Effects of random particle distribution (*a*) CSC field for simulated trajectories with temporal discretisation, initial error and path error and a correct value of $\epsilon(3) = 0.3$. (*b*) Normalised CSC field error; i.e., (2.10) versus candidate parameter $\epsilon(3)$. Error metric is normalised by the lowest error over the range of $\epsilon(3)$ tested. Solid black line indicates the error averaged over 10 individual sweeps, and shaded region bounds the range of errors seen for an individual sweep. Red dotted line indicates the correct value for the parameter $\epsilon(3)$.

is unable to detect the kinematic dissimilarity between the vortices and the jet, and instead identifies as most dissimilar the drifters in the jet from those directly outside the jet. Due to this, the CSC field error is unable to correctly identify the model parameter being analysed. This highlights the importance of understanding the effects driving real and simulated trajectories apart, and how to mitigate these effects on the CSC field for simulated particles by using appropriate drifter placement.

It is also necessary to consider the relative importance of the individual sources of error considered in this analysis. To this end, figure 12(*a*) shows the CSC field error for simulated trajectories without both temporal discretisation and stochastic position error. In this case, the trajectories are deterministic, and the error at the correct parameter value of $\epsilon(3) = 0.3$ is identically zero, because the simulated trajectories perfectly match the real set of trajectories. Figure 12(*b*) shows the CSC field error with temporal discretisation but no initial or path error and figure 12(*c*)

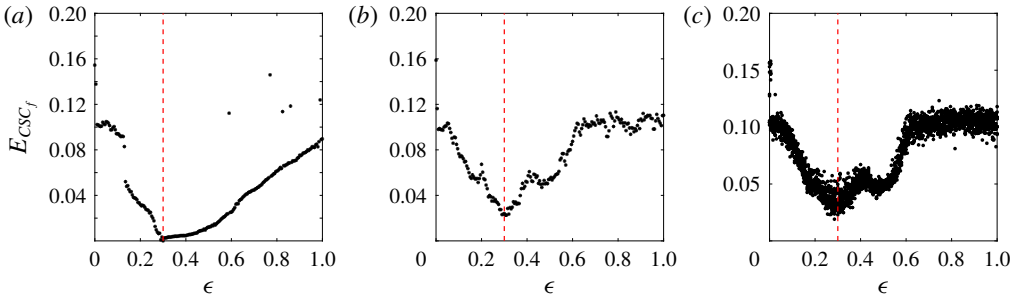

FIGURE 12. (Colour online) CSC field error; i.e., (2.10) versus candidate parameter $\epsilon(3)$. Error metric is not normalised in this case. Red dotted line indicates the correct value for the parameter $\epsilon(3)$. (*a*) Trajectories advected with no temporal discretisation and no path or initial error. (*b*) Trajectories advected with temporal discretisation and no path or initial error. (*c*) Trajectories advected with temporal discretisation, path error and initial error, data collected over 10 independent parameter sweeps.

includes both temporal discretisation and stochastic error. When considering only temporal discretisation, the trajectories are deterministic given a value of the unknown parameter $\epsilon(3)$. The CSC field error for temporal discretisation has a global minimum at the correct value of $\epsilon(3) = 0.3$, but the error values are generally higher, and the shape of the error curve is different, exhibiting several local minima. The difference between figures 12(*b*) and 12(*c*) shows that the addition of initial and path error to temporal discretisation serves to increase the noise in the error curve, which leads to the spread of identified parameter values around the correct value, as discussed previously. It is important to note that while the parameters chosen for temporal discretisation and position uncertainty do lead to some error in the identification of the test parameter, the CSC field error still provides a better metric for identifying unknown parameters than the particle displacement error.

4. Conclusions

The two test cases examined in this study identify and characterise the influence of initialisation errors, temporal discretisation and stochastic path error on the process of Lagrangian data assimilation for model parameter estimation. By considering a field-based definition of the underlying flow structure using coherent structure colouring, model parameters can be more accurately and robustly determined than by considering particle displacement errors alone. This study highlights the value of coherent structure identification, and the CSC algorithm in particular, in model parameter estimation. This method can potentially be extended from the determination of model parameters in analytical flows to the more complex problem of Lagrangian data assimilation into large-scale oceanic and atmospheric models. The CSC-based method could eliminate the need for simulation ensembles in order to accurately estimate model parameters, thereby advancing the state of the art in numerical flow prediction.

Acknowledgements

This work was supported by the U.S. National Science Foundation and by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

REFERENCES

APTE, A., JONES, C. & STUART, A. 2008 A Bayesian approach to Lagrangian data assimilation. *Tellus* A **60** (2), 336–347.

BÖNING, C. & HERMANN, P. 1994 Annual cycle of poleward heat transport in the ocean: results from high-resolution modeling of the North and Equatorial Atlantic. *J. Phys. Oceanogr.* **24**, 91–107.

FROYLAND, G. & PADBERG-GEHLE, K. 2015 A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data. *Chaos* **25**, 087406.

GUCKENHEIMER, J. & HOLMES, P. 1983 *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer.

HADJIGHASEM, A., KARRASCH, D., TERAMOTO, H. & HALLER, G. 2016 Spectral-clustering approach to Lagrangian vortex detection. *Phys. Rev.* E **93**, 063107.

HALLER, G. 2000 Finding finite-time invariant manifolds in two-dimensional velocity fields. *Chaos* **10** (1), 99–108.

IOUALALEN, M., ASAVANANT, J., KAEWBANJAK, N., GRILLI, S., KIRBY, J. & WATTS, P. 2007 Modeling the 26 December 2004 Indian Ocean tusnami: case study of impact in Thailand. *J. Geophys. Res.* **112**, C07024.

KALNAY, E. 2002 *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.

MACLEAN, J., SANTITISSADEEKORN, N. & JONES, K. 2017 A coherent structure approach for parameter estimation in Lagrangian data assimilation. *Physica* D **360**, 36–45.

MARIANO, A., KOURAFALOU, V., SRINIVASAN, A., KANG, H., HALLIWELL, G., RYAN, E. & ROFFER, M. 2011 On the modeling of the 2010 Gulf of Mexico oil spill. *Dyn. Atmos. Oceans* **52** (1–2), 322–340.

OLIVIERI, S., PICANO, F., SARDINA, G., IUDICONE, D. & BRANDT, L. 2014 The effect of the Basset history force on particle clustering in homogeneous and isotropic turbulence. *Phys. Fluids* **26**, 041704.

PUTMAN, N. & HE, R. 2013 Tracking the long-distance dispersal of marine organisms: sensitivity to ocean model resolution. *J. R. Soc. Interface* **10**, 20120979.

RYPINA, I., BROWN, M. & BERON-VERA, F. 2007 On the Lagrangian dynamics of atmospheric zonal jets and the permeability of the stratospheric polar vortex. *J. Atmos. Sci.* **64**, 3595–3610.

SCHLUETER-KUCK, K. & DABIRI, J. 2017*a* Coherent structure colouring: identification of coherent structures from sparse data using graph theory. *J. Fluid Mech.* **811**, 468–486.

SCHLUETER-KUCK, K. & DABIRI, J. 2017*b* Identification of individual coherent sets associated with flow trajectories using coherent structure coloring. *Chaos* **27**, 091101.

SHADDEN, S., LEKIEN, F. & MARSDEN, J. 2005 Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica* D **212** (3–4), 271–304.

THIFFEAULT, J. 2010 Braids of entangled particle trajectories. *Chaos* **20**, 017516.

YAMAZAKI, Y., LAY, T., CHEUNG, K., YUE, H. & KANAMORI, H. 2011 Modeling near-field tsunami observations to imrpove finite-fault slip models for the 11 March 2011 Tohoku earthquake. *Geophys. Res. Lett.* **38**, L00G15.